中医诊疗知识库术语规范化研究

■ 马捷[1,2] 吴文玲[1] 孙恒宇[3] 王春秀[1]

[1] 吉林大学管理学院 长春 130022 [2]吉林大学信息资源研究中心 长春 130022

[3] 吉林省中医药科学院第一临床医院 长春 130022

摘 要：[目的/意义]中医术语是中医知识的载体，是构成中医诊疗知识库的重要组成部分，面向中医诊疗知识库的术语规范化研究，有利于提高诊疗知识库医案数据信息的质量，弘扬中医文化。[方法/过程]确定中医诊疗知识库术语提取及选词范围，构建术语规范档结构，并针对术语规范中存在的问题提出词义控制策略，结合具体实例说明知识库术语提取及规范档建立的过程。[结果/结论]基于中医诊疗知识库的术语规范化工作，可以提高中医诊疗知识库医案的利用效率，推进中医现代化与国际化进程。

关键词：中医诊疗知识库 中医术语 术语规范化

分类号：G251.2

DOI：10.13266/j.issn.0252-3116.2021.02.002

1 引言

随着信息化时代的到来，互联网和大数据技术为中医的发展带来了很好的机遇。在发布的《中医药发展战略规划纲要》中，国家明确做出如下指示：在2020年之前要达成公民都能享有基础医疗服务的任务，大力推进互联网与中医相结合，构建线上远程医疗的新模式[1]。知识库是一种提供知识服务的人工智能系统[2]，已在多个领域有了广泛的应用。通过构建中医诊疗的知识库可以对中医诊疗过程中的中医知识和经验进行存储，为中医数据共享与信息交流、利用提供基础保障，智能化辅助医生进行诊疗，提高中医诊疗水平，同时也对于中医文化的继承与发展具有非常大的作用。目前已有了这类的产品，如医脉通诊疗知识库、好大夫在线、医护网及腾讯医典等在线诊疗网站。但是在中医知识库资源的组织过程中，存在诸多中医术语不规范的问题，多词一义、一词多义、用词习惯不统一、病名生僻难懂及术语间概念关系不明确的现象常有发生，这些问题给中医诊疗知识的分析和挖掘带来造成了阻碍，也影响了机器处理和机器理解的准确性和效率。目前，关于中医术语规范化的研究已经有了一定的成果，但是专门面向诊疗过程知识库构建的研究还非常罕见。笔者基于中医诊疗知识库，确立了术语提取方式及选词范围，从基本结构、格式规范、类别及其与标准的对应关系4个方面构建术语规范档结构，针对术语规范中存在的问题提出词义控制策略，并结合具体实例说明知识库术语提取及规范档建立的过程。中医术语是中医知识的载体，是构成中医诊疗知识库的重要组成部分，是诊疗知识库中医案得到认可与理解的先决条件。中医术语的规范化研究是进行机器理解的关键前提，对于构建中医诊疗知识库而言是一项极为重要、任重而道远的基础性工程，它不仅可以促进中医行业学术的交流与思考，还对于中医学知识的传承与发展、推动中医现代化与国际化具有深远意义。目前网络信息资源逐渐丰富，自媒体发展日新月异，中医诊疗医案的网络共享，都对中医诊疗方面的术语标准化、规范化提出了迫切的要求。

2 文献回顾

如将医案的结构及要素比作医案的“骨架”，则中医术语为医案的“血肉”。中医术语有其独特性，例如《伤寒论》《神农本草经》这样的经典中医著作，文言色彩浓重，加之我国疆土广阔、各民族间的方言众多，导致大部分术语具有笼统、模糊的特性。朱建平[3]将中医术语特点概括为历史性、人文性、用具体名词表达抽象概念性、定性描述性等特点。中医作为国粹之一，逐

作者简介：马捷(ORCID:0000-0002-1471-2143)，教授，博士，博士生导师；吴文玲(ORCID:0000-0001-7396-7261)，硕士研究生，通讯作者，E-mail:2081364101@qq.com；孙恒宇(ORCID:0000-0003-0882-7507)，主治医师；王春秀(ORCID:0000-0002-5632-0519)，硕士研究生。

收稿日期:2020-05-19 修回日期:2020-07-31 本文起止页码:17-26 本文责任编辑:徐健

渐走出国门,不规范的术语成为阻碍中医药登上世界学术舞台的重要因素之一[4]。

针对上述问题,科研工作者在术语名词规范化方面进行了大量工作。首先,颁布了中医药名词术语国家标准、行业标准及国际标准。其次,借助现代信息技术建立了术语数据库。中国中医科学院研究院开发了一系列中医数据库,总量达 45 个,包含了各类标准化数据,提供数十万条标准化术语内容,极大提高了术语的标准化、信息化程度[5]。另外,针对术语标准应用度不高的问题,郭玉峰等[6]提出在真实世界下中医术语规范化应用技术体系的研究思路,指出术语标准在实践科研中应用程度不高的主要原因在于缺乏与标准配套的规范化应用模式与工具,为解决术语标准化的问题,需要通过运用信息化技术与标准条例来规范调用现有标准化术语;倪菲等[7]运用文献计量学的研究方法对中医脾胃系统疾病范围的症状术语进行标准化研究;李灵巧、包素珍[8]对中医术语标准化的意义进行了阐明,说明了其对文化传承、学术交流及外译的重要作用,并提出中医术语标准化应在保留自身特点及优势的基础上,通过现代化的思维及技术方法来实现;朱建平[9]回顾 2000 年以来中医术语规范化的成果和工作进展,指出目前制约中医现代化发展的重要因素之一是中医术语的标准化问题,并对于现下的中医术语问题,提出了相应的建议和展望;许吉等[10]对颁布的几个中医术语国家标准从总体情况、用词习惯、收词范围以及展现形式的方面进行了比较分析;王传池等[11]对近 5 年中医药标准化的相关政策、新闻报道从中医药标准化的现状及存在问题的角度进行解读。

综上所述,众多学者均提及中医术语标准化对中医发展的重要作用,指出中医术语标准化现存问题并提出了相应的建议。近年来,随着信息技术的发展,知识库已在医学领域有了广泛的应用。虽有专家发表中医知识库的相关文献,但尚未有学者对中医诊疗知识库中的术语规范化工作做出研究。鉴于此,笔者以中医诊疗知识库为研究对象,对知识库中术语的选词范围、规范档结构、存在问题等相关规范化工作进行探讨并举出实例为证,以期促进中医术语标准化研究的多元化发展。

3 中医诊疗知识库的术语提取及选词范围

3.1 以现有国家、行业标准为基本依据

自 20 世纪 90 年代以来,中医术语标准化工作就受到了国家和行业部门的重视与支持,相继开展了一系列中医行业术语规范化标准的研究工作,颁布了几部中医术语标准,中医诊疗知识库应首先以此为基本依据。目前已颁布的中医诊疗相关国家标准有:《GB/T 20348 – 2006 中医基础理论术语》《GB/T 16751.3 – 1997 中医临床诊疗术语 – 治法部分》《GB/T 16751.2 – 1997 中医临床诊疗术语 – 证候部分》《GB/T 16751.1 – 1997 中医临床诊疗术语 – 疾病部分》《GB/T 15657 – 1995 中医病症分类及代码》。行业标准有:《ZY/T 001.1 – 94 中医内科病症诊断疗效标准》《ZY/T 001.2 – 94 中医外科病症诊断疗效标准》《ZY/T 001.3 – 94 中医妇科病证诊断疗效标准》《ZY/T 001.4 – 94 中医儿科病证诊断疗效标准》《ZY/T 001.5 – 94 中医眼科病证诊断疗效标准》《ZY/T 001.6 – 94 中医耳鼻喉科病证诊断疗效标准》《ZY/T 001.7 – 94 中医肛肠科病证诊断疗效标准》《ZY/T 001.8 – 94 中医皮肤科病证诊断疗效标准》《ZY/T 001.9 – 94 中医骨伤科病证诊断疗效标准》,如表 1 所示:

表 1 中医诊疗术语国家及行业标准

类别	名称	标准编号	生效时间	制定方
国家标准	中医基础理论术语	GB/T 20348 – 2006	2006 – 10 – 01	国家质量监督检验检疫
	中医临床诊疗术语 – 治法部分	GB/T 16751.3 – 1997	1997 – 10 – 01	国家技术监督局
	中医临床诊疗术语 – 证候部分	GB/T 16751.2 – 1997	1997 – 10 – 01	国家技术监督局
	中医临床诊疗术语 – 疾病部分	GB/T 16751.1 – 1997	1997 – 10 – 01	国家技术监督局
	中医病症分类及代码	GB/T 15657 – 1995	1996 – 01 – 01	国家技术监督局
行业标准	中医内科病症诊断疗效标准	ZY/T 001.1 – 94	1995 – 01 – 01	国家中医药管理局
	中医外科病症诊断疗效标准	ZY/T 001.2 – 94	1995 – 01 – 01	国家中医药管理局
	中医妇科病证诊断疗效标准	ZY/T 001.3 – 94	1995 – 01 – 01	国家中医药管理局
	中医儿科病证诊断疗效标准	ZY/T 001.4 – 94	1995 – 01 – 01	国家中医药管理局
	中医眼科病证诊断疗效标准	ZY/T 001.5 – 94	1995 – 01 – 01	国家中医药管理局
	中医耳鼻喉科病证诊断疗效标准	ZY/T 001.6 – 94	1995 – 01 – 01	国家中医药管理局
	中医肛肠科病证诊断疗效标准	ZY/T 001.7 – 94	1995 – 01 – 01	国家中医药管理局
	中医皮肤科病证诊断疗效标准	ZY/T 001.8 – 94	1995 – 01 – 01	国家中医药管理局
	中医骨伤科病证诊断疗效标准	ZY/T 001.9 – 94	1995 – 01 – 01	国家中医药管理局

3.2 以中医常用现代文用语为补充

实践表明,现有的中医术语国家标准与行业标准无法覆盖所有中医诊疗科研工作中所需术语。在中医诊疗实际操作中,所需的术语不仅来自于一些固定的、标准化的术语,更多的是来源于实际工作、海量临床诊疗以及科学研究的术语,因此除了将中医术语国家标准与行业标准为基本依据外,还需要以中医常用的现代文用语作为补充,建立用来实现不同功能目标、各有针对性、在收词量有所互补的、多来源的术语集。

对于以上标准中没有覆盖但是实际操作中所需要的名词术语,可以参照已发布并且受到广泛认可、具有一定学术水准的中医词典和中医教科书,如《中医名词词典》《中医词库》《中医药常用名词术语辞典》《中医药词典》《中医大辞典》《中医药学主题词表》《中医基础理论》《中医诊断学》等书籍。此外,还可以从知名度高,有影响力的中医期刊、中医名医名著及在线医疗平台上抽取筛选整理出合适的中医名词术语对中医诊疗知识库名词术语集进行及时动态的更新补充。为了保证中医诊疗知识库中名词术语的专业性及标准性,这些补充录入的中医名词术语必须与前期已经录入的术语保持一致的原则且保证其规范性,只有通过中医诊疗术语集的标准化认证机制,以及咨询领域专家意见之后,才能正式录入进中医诊疗知识库术语集中。

4 中医诊疗知识库术语规范档结构

4.1 中医诊疗知识库基本结构

中医诊疗知识库系统的基本结构主要分为 3 个层面:基础数据层、规则标准层及诊疗方案推理层。具体如图 1 所示:

基础数据层
· 医案数据库 中医药知识库 中医文献库

规则标准层
· 中医药学术语:《中华人民共和国药典》《中国药名通用名称》
· 中医诊疗术语:国家及行业现行标准
《术语工作 原则与方法(GB/T 10112-1999)》
《术语工作 概念和术语的协调》

诊疗方案推理层
· 倒排索引表 相似度 贝叶斯分类算法

图 1 中医诊疗知识库基本结构

基础数据层包括医案数据库、中医药知识库、中医文献库。医案数据库以海量中医诊疗病例为基础,医案资料来源于真实世界实际临床诊疗中的病人电子病历,得到病人认可及隐去私人信息,将医案字词按照中医诊疗知识库术语规范进行修改后,录入知识库系统;中医药知识库包含疾病知识详解、中药知识详解、中医治法知识详解和古代经典名方,用户可以通过学科分类及主题检索或者页面的知识标引[12],快速方便学习及查询相关中医诊疗知识;中医文献库则以万方、知网中医文献为基础,由专家进行审核及跟踪负责,根据每个词条在国内外的最新研究前沿,对相关文献进行实时更新,提供最新可靠的中医诊疗文献资源,用户可享有在线阅读和收藏下载文献全文的权限,为临床诊疗决策提供依据,达到学习和辅助决策的目的。

中医诊疗知识库中的规则标准层主要指的是术语规范方面。中医药学术语以国家药典委员会编写的最新版《中华人民共和国药典》和《中国药名通用名称》为准,不使用药物的商品名;中医诊疗术语则以国家及行业现行标准为基本依据,后补充的术语应符合国家颁布的《术语工作 原则与方法(GB/T 10112－1999)》及《术语工作 概念和术语的协调》相关标准。

诊疗方案推理层是通过将诊疗知识库的病症信息分词,设立倒排索引表,计算所查询的症状特征与诊疗知识库相关疾病内容的相似度,根据相似度排序,诊断疾病类别。再通过用户的反馈信息,进一步优化诊断,使用贝叶斯分类算法进行诊疗方案的推荐[13]。

4.2 中医诊疗知识库术语格式规范

中医诊疗知识库术语的数据类目应必须包含条目编号、术语名称及定义,其余按照需要可增加对应词、按语、文献汇考等补充信息。

条目编号格式采用两层的结构。第一层为总的术语顺序编号,置于条目编号的顶格列,可反映诊疗知识库中术语的总量;第二层为术语的代码编号,由英文字母加阿拉伯数字的方式组合构成,字母代表术语大类的标识符,数字则为标识符下的顺序编号,代码编号以此来反映术语概念之间的层次及关系,如咳嗽病的代码编号为 BNF010,其下级病症外感咳嗽病标记为 BNF011,以此类推。

术语名称为了以防术语之间的混乱,秉持唯一性原则,只采用一个正名,应该简明而准确,置于条目编号之后。优先采用中医术语国家标准和行业标准的术语名称,若出现一义多词的情况,另一词可标记为同义词。

定义置于术语名称之后,另起一行。定义的表述需避免重复和矛盾,不可使用"这""此""该"这样的指示代词作为定义的开头。一般而言,术语定义要包含该术语的上位概念,并且说明它与其他同级概念的区

别特征。若有引用,可以在定义之后用方括号括起,标注该定义的来源出处[14]。

其他补充信息,如对应词指的是与术语意思相近或相反的词;按语是对术语定义进行必要的补充说明;文献汇考则列举的是与术语相关的文献来源。这些补充信息均置于定义之后,根据需要进行增补。

4.3 中医诊疗知识库术语规范档类别

中医主要是通过"四诊",即望、闻、听、切来诊断疾病,医生再根据具体病症决定采取何种治疗方法以及服用何种药物进行治疗。中医诊疗知识库术语规范档的类别应包括病症名称、症状表现、治疗方式、药品名称、疗效标准。

4.4 术语规范档类别与标准的对应关系

术语规范档类别与中医术语的国家标准与行业标准相对应。病症名称对应国家标准《GB/T 15657 – 1995 中医病症分类及代码》;症状表现对应《GB/T 16751.2 – 1997 中医临床诊疗术语 – 证候部分》和《GB/T 16751.1 – 1997 中医临床诊疗术语 – 疾病部分》;治疗方式对应《GB/T 16751.3 – 1997 中医临床诊疗术语 – 治法部分》;药品名称对应《中华人民共和国药典》和《中国药名通用名称》;疗效标准对应中医行业标准《ZY/T 001.1 – 94 中医内科病症诊断疗效标准》《ZY/T 001.2 – 94 中医外科病症诊断疗效标准》《ZY/T 001.3 – 94 中医妇科病证诊断疗效标准》《ZY/T 001.4 – 94 中医儿科病证诊断疗效标准》《ZY/T 001.5 – 94 中医眼科病证诊断疗效标准》《ZY/T 001.6 – 94 中医耳鼻喉科病证诊断疗效标准》《ZY/T 001.7 – 94 中医肛肠科病证诊断疗效标准》《ZY/T 001.8 – 94 中医皮肤科病证诊断疗效标准》《ZY/T 001.9 – 94 中医骨伤科病证诊断疗效标准》,如图 2 所示:

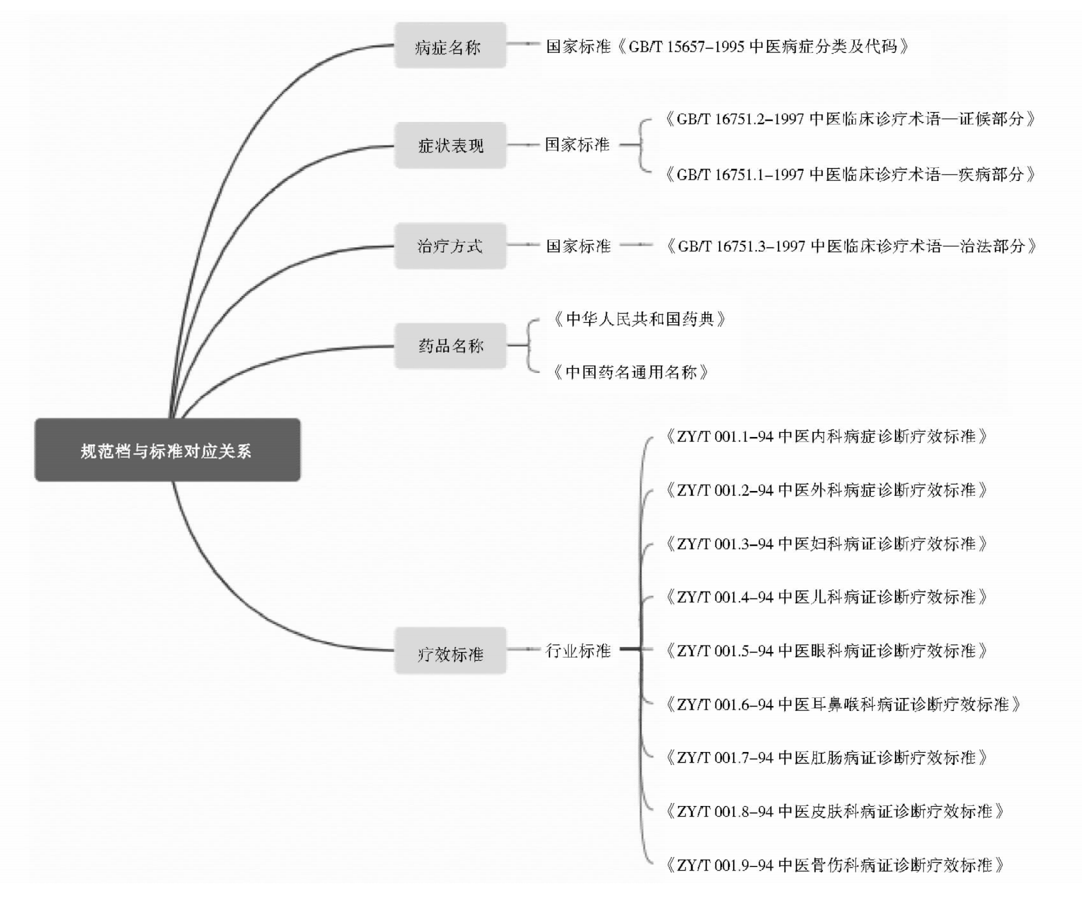


图 2 规范档与标准对应关系

5 中医诊疗知识库术语规范化存在的问题及词义控制策略

5.1 多词一义

中医术语分布于各类标准性文件、期刊、教材、词典及书籍中,数量庞大。如果从其中的单个资料来看,术语之间没有什么严重问题。但是当将多种来源及各自标准不同的术语集中起来,可能产生术语之间语义的冲突、交叉或重叠,造成一义多词的现象,问题便接踵而至,具体表现在同一个概念在不同的文章中可能

会用不同的术语来表示[15]。

多词一义的情况，举一个中医眼科疾病的例子说明，表示由于痰液阻滞在体内，气血不通，上火所造成的，症状为眼球突出，眼白处充血，红如鸪眼的内、外障疾病，中医临床诊疗术语疾病部分采用"鸪眼凝睛"的名称[16]，而中医眼科病证诊断疗效标准则用"鹘眼凝睛"来表示这一疾病[17]。实际上"鸪"与"鹘"是两种不同的鸟类，但"鸪眼凝睛"与"鹘眼凝睛"在中医领域表示的却是相同的意思。多词一义的存在毫无疑问会使术语的使用者造成选择上的困难，带来巨大的混乱。

一个关键的传统术语遵循的原则是，一个特定的概念仅可以由一个术语来表示，即术语具有唯一性。针对以上情况，应对的策略是从不同资料中提取出所有表示同一含义的词，然后使用分词软件对它们进行词频统计，如果词频相差较大，那么选择在所有资料来源中出现频次最高的术语，作为表示中医诊疗知识库中该概念的唯一规范正式术语[18-19]。此外，术语描述的目标是实现术语规范化，这意味着选择术语并将其输入数据库的目的是推荐对特定概念使用最可接受的术语，应将其他表示同一概念的术语另外单独列出，但对每一个同义词赋予一定的规范地位：承认的术语、不推荐的术语、过时的术语。

5.2 一词多义

中医学有着悠久的历史，它是从数千年来古代人民与疾病斗争的实践经验中经过不断总结归纳，形成的医学理论体系。加之，中国疆土广阔，民族数量多达56个，受方言的局限性及地区的闭塞性，信息交流不便，因此不可避免地导致了中医术语一词多义的问题。举例说明中医术语中一词多义的现象，如"命门"一词，就有3种含义，既代表人体的穴位名称，即督脉命门穴，它的位置在肚脐的正对面腰椎处；又可以表示为人的组织器官，在五脏学说中指肾脏，在《黄帝内经》中指眼睛；同时还象征着人生命和元气的根本[20]。此外，对于中医术语的"气"可以理解为"能量""空气""物质""活动""症象""精神状态"6种不同的理解。

中医学者许志泉曾统计了《中医大词典-基础理论分册》中的医学术语，发现此书中的中医术语一词多义出现的比率竟然达到了15.8%，大大超出了西医学0.3%的比率[21]。一词多义是违反了标准术语准则的"同一律"原则，造成中医术语概念的模糊不清，使得中医的学习者需要花更多的时间来辨明词义，无形中给中医学习增大了难度，还严重制约了中医学术的交流，阻碍中医学的继承与发展，是中医术语标准化的一大绊脚石。

针对中医术语的一词多义现象，可以先将术语的多种定义都罗列出，删除不常用的定义，再将不同义项进行分化，通过在术语名称后加上标注的方式，使得术语名词能够对应得上各个不同的义项。对于难以分化的一词多义问题，可以采用德尔菲法，这是一种有效的专家咨询法，已广泛运用于中医领域。首先由业内具有一定水准的学者、名医组成中医领域的专家小组，之后根据问题编制咨询表，进行分轮多次匿名的咨询调查，再将每次收集到的意见不断整理、归纳与统计，最终得到结果，采用其受到普遍认可、通俗易懂、科学合理的定义[22]。例如，"喉窝"一词，它原本有两个含义，一是指人身体位于胸骨上的隐窝处，二是指在咽腭弓和舌腭弓之间的凹陷处，经专家研究讨论决定后，取其第二个意思作为"喉窝"的中医定义，摈弃胸骨上隐窝处的含义[23]。

5.3 用词习惯不统一

多个术语标准之间存在用词习惯不统一的现象，包括病名后缀使用习惯不同、选词不同及用字不同的问题。这些问题使得中医术语的使用无法统一，制约了术语标准化的进程。

(1)病名后缀。不同术语标准之间常常有病名后缀是否加"病"字或者"证"字的用词习惯不统一的现象。以同为中医术语国家标准的《GB/T 15657-1995中医病症分类及代码》和《GB/T 16751.1-1997中医临床诊疗术语-疾病部分》为例，《中医病症分类及代码》均会在疾病名之后加上"病"的后缀，而《中医临床诊疗术语-疾病部分》则有的病名后缀没有加"病"字，直接以病名命名。如，《中医病症分类及代码》中的肺衰病、肺痿病、胸痹心痛病，在《中医临床诊疗术语-疾病部分》标准中直接以肺衰、肺痿、胸痹心痛表示。

(2)用字。中医在开方时，由于有病人在等待，医生为了提高速度，养成了潦草快写的书写习惯，常常按照个人习惯用简化字或者同音字代替，长此以往造成了中医术语标准之间用字习惯不同的问题。比如，在《中医临床诊疗术语》中的"乳瘘""鼓胀""疤痕""熄风"，在《中医药学名词》及《中医病症分类及代码》中分别对应的是"乳漏""臌胀""瘢痕""息风"[10]。

中医术语用词习惯不统一的问题，亟待进一步规范化。首先，需要加强医生及中医学者的职业道德和学术素养教育，使他们充分意识到使用不规范术语所带来的混乱和损失，改正随意胡乱使用术语的习惯，熟悉和掌握规范化中医术语的表达，养成良好的习惯。

另外，对于上述类似中医药国家、行业标准彼此之间术语用词习惯不同，建议中医药相关部门对已颁布的标准定期进行修订，对术语名词做统一规范化处理，只采用同一个术语名词形式表示，避免出现同一术语名词两种不同表示方法的现象。中医诊疗知识库的术语用词习惯首先参照国家标准和行业标准，若存在冲突，再以《中国中医药学主题词表》中的主题词作为依据。此外，对于尚未收录进已有国家、行业标准及《中国中医药学主题词表》中的无规范化中医名词术语，应确保其在中医诊疗知识库中相统一，严格执行一个标准，切勿一个术语多个表达。

5.4 病名生僻难懂

中医自古时候流传下来，文字上带有古代汉语的特点，有些中医术语是字词比较生僻，甚至连如何发音也不知，如“瘰疬”“癥瘕”“瘛瘲”“龂”“燠暍”等。而有些中医术语则是单个字词简单，但是组合起来的字义与病症看似联系不多，难以理解其内涵定义。比如：“伏兔疽”，多指由胃火导致的，生于大腿的毒疮；“正水”，指风热湿毒所致，表现为水肿、高血压、尿少的疾病；“肥气”，指体内淤血内积，腹部肿胀、出血的积聚型疾病；“伏梁”，指由肠道秽浊所致，主要表现为气血阻滞，腹部不适，右下腹包块的疾病。

病名生僻难懂，会影响人们对于中医学术知识的理解与交流，阻碍中医走出国门。试想如果这些中医术语连许多医学生甚至是诊疗医生都无法理解，那他们该如何向病人解释，病人和社会如何能明白，中医如何走向世界呢？这些生僻难懂的中医术语不仅不能体现疾病的病因、病状及特点，反而给人们造成了理解上的障碍，更别提如何利用这些中医术语去指导和分析病情。

考虑到中医术语与一般语言不同，中医术语既涉及中医专业知识体系又牵扯到语言系统，因此在处理生僻难懂的病名术语时，应充分考虑到语言及专业的综合因素[24]。中医术语标准化工作需要中医学者与语言学家的相互合作交流，在深刻理解术语的内在含义及特点的基础上，与时俱进，适当地对字面生僻难懂的中医术语赋予新词，保证术语的准确性及简明性，力求简单与清楚。中医不能一成不变，以孤立静止的态度看待中医，必须对晦涩难懂的中医术语进行改造，在继承中发展中医，使中医不断现代化，被更多人所理解与接受，扩大中医的影响力。

5.5 术语间的概念及关系不明确

古时候受到客观条件的限制，中医诊断病情的方式有限，主要通过望、闻、问、切的方式，然而面对病症之间的概念与关系比较困难的情况，会出现术语概念、层次混淆的问题。如，人们对“证”“症”“征”“证候”这几个术语的认识一直存在差异，常常出现这些术语的混用，同时还有位于列表前的术语又平行列入其下一级术语中的情况出现。而这些对于术语概念认识的不同将导致中医学者难以达成共识，直接影响后期具体病症的命名、分类等标准化工作以及学术之间的交流。

概念应具备概念与特征、特性之间的关系，即本体知识体系中的属性[25]。一个术语的内涵是反映其本质属性的总和，而外延是该术语概念所反映的那一个类，也就是说该术语概念都适用于哪些范围。因此确定一个中医术语的概念，需要界定清楚该术语的内涵与外延，构建一个清晰的概念系统，从而从源头对这些术语加以区别。构建一个连贯的中医术语概念系统，需要将中医诊疗数据库中的所有中医术语都按照上下级的明确关系进行排列，表明术语间的从属逻辑关系，以便它们形成一个清晰的层次系统。以中医疾病术语“咳嗽病”为例，其内涵为由外感等因素所致，肺失宣肃，气冲呼吸道，发出咳声的一种病症。其下位词为外感咳嗽病和内伤咳嗽病，受外感诱因的是外感咳嗽病，而因肺脏功能失调导致的为内伤咳嗽病。咳嗽病的同位词有肺衰病、肺漏病、喘病等，它们的上位词是肺系病类，同属于出现在肺部的疾病，如图 3 所示：

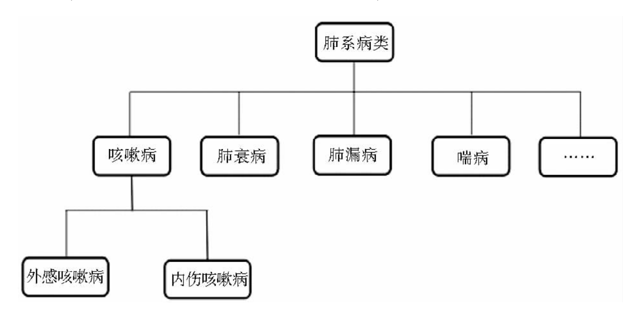


图 3 肺系病类概念体系

6 中医诊疗知识库术语提取及规范档建立方法与实例分析

6.1 多词一义

6.1.1 建立方法与流程

中医知识本身具有一定的模糊性及意会性，受历史因素的影响，古时候人们对于中医概念的认识并不具体，只能通过身边的实际病例去体会与想象，这也造成了每个人在中医的理解上可能存在差异，出现多词

一义的现象。中医术语中多见多词一义，即同义词的问题，这会造成术语使用与理解上混乱的情况，不利于术语规范化。构建中医诊疗术语规范档，必须对此问题加以解决，秉持术语规范中的同一律原则，只选取一个术语作为中医诊疗知识库中的正式规范用语。多词一义情况下，中医诊疗知识库规范档的建立流程如图4所示：

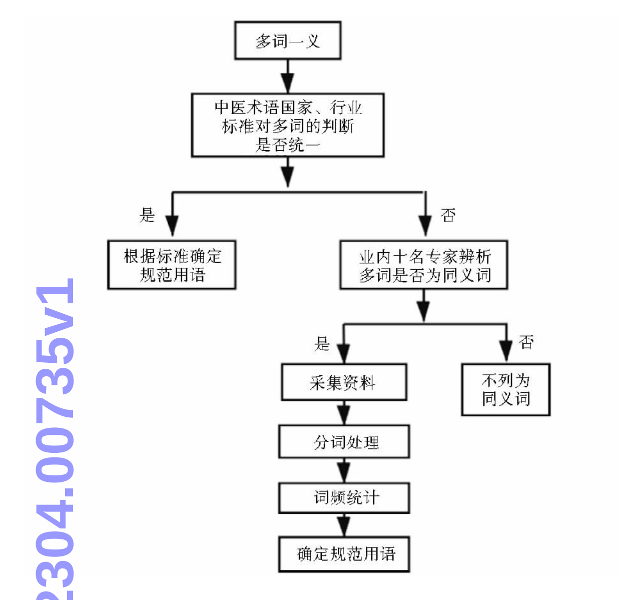


图4　多词一义中医诊疗知识库术语规范档建立流程

首先，根据上述中医诊疗知识库的术语提取及选词范围，其规范档的建立应先以现有中医术语国家标准、行业标准作为基本依据。在多词一义的情况下，若中医术语国家、行业标准对于多词中的标准术语判断统一，则可以直接将该术语确定为中医诊疗知识库的规范术语，其他术语标记为不推荐术语或者过时术语。其次，由于多词一义相较于其他的问题而言，它的操作流程比较困难。在处理多词一义问题的时候，其中一个重要的步骤是需要判断多词是否表达的是同一个意思，在此基础上才能继续进行下去。因此对于那些无法以中医术语国家标准、行业标准作为依据的中医术语（如多个标准之间关于多词是否为同义词存在争议，以及尚未收录进中医术语国家标准、行业标准中的中医常用现代用语的情况），需要由中医行业内的十名专家对多词进行辨析，若大多数专家认为构成同义词，可以采用分词处理后统计词频的方式，将在中医核心期刊中被使用频次较高的术语作为中医诊疗知识库的规范术语；若意见分歧较大，多数专家认为多词之间存在较大差异，不能构成同义词，则多词不做同义词处理，列为独立的两个术语。

在此，本文以中医术语“营气”“荣气”和“湿热蕴脾”“脾经湿热”为例具体说明多词一义情况下如何抽取中医术语及建立规范档。

6.1.2　术语提取及规范档建立——以“营气”“荣气”为例

百度百科、医学百科对于“荣气”词条的标注为：荣气即营气、荣气的意思是营气，将“营气”与“荣气”视为同义词，皆指循行于脉中而具有营养作用的气。“营气”与“荣气”常常出现在不少中医期刊、论文中，混用的情况至今时有发生。在中国知网上主题检索“营气”词条，共出现350条结果，文章中含有“营气”这一术语的中医文献最新发表时间是2020年7月；检索“荣气”词条，则有25条结果，且文章中含有“荣气”这一术语中医文献的最新发表时间是在2019年4月。

根据上文的中医诊疗知识库术语提取原则，首先应该以中医术语国家标准、行业标准为基本依据，若没有再根据它们在中医核心期刊中出现的频次确定中医诊疗知识库正式用词。经过检索，“营气”在中医术语国家标准中共出现16次，“荣气”出现0次，故将“营气”作为中医诊疗知识库的正式用语，将“荣气”标记为不推荐或被取代术语。

6.1.3　术语提取及规范档建立——以“湿热蕴脾”“脾经湿热”为例

在《GB/T 16751.2－1997 中医临床诊疗术语－证候部分》中“湿热蕴脾”和“脾经湿热”为同义词，同指由于湿热阻滞，脾失健运，导致腹胀，呕恶纳呆，肢体困重，便溏不爽，或面目发黄，身热不扬，汗出热不解，渴不多饮，舌红苔黄腻，脉濡数的湿热证候[26]，其中“湿热蕴脾”是作为主题词，“脾经湿热”列为了“湿热蕴脾”的同义词。但是在《GB/T 15657－1995 中医病症分类及代码》中，“脾经湿热”则是主题词，没有出现“湿热蕴脾”这一中医术语。二者相较，两部中医术语国家标准中制定的标准不一，在这种情况下，显然已无法将中医术语国家标准作为确定“湿热蕴脾”与“脾经湿热”何者为中医诊疗知识库规范术语的依据。因而，进入下一步，由中医行业内的十名专家对“湿热蕴脾”和“脾经湿热”进行辨析，判断它们是否为同义词。笔者咨询了自于北京同仁医院传统中医科、北京中医学院、深圳市中医院肛肠科、吉林省中医药科学院第一临床医院胃肠科、长春市中医院泌尿科、吉林省中医药科学院肺病科、江西省中医院、江西湖口妇幼保健院中医中药科、辽宁中医药大学附属医院、山东省淄博市市立医院中医科等的十多名临床中医专家。由于专家对于“湿热蕴脾”和“脾经湿热”是否为同义词的争议较大，

并不能确定二者为同义词，因此在中医诊疗知识库中"湿热蕴脾"和"脾经湿热"不做同义词处理，列为两个独立的术语。

6.2 一词多义

6.2.1 建立方法与流程

中华文化博大精深，汉字蕴意丰富，一词多义的情况常有发生，中医学发展历史悠久，也不例外。中医术语中的一词多义，可能是同一术语在不同领域存在不同的含义，如"麻木"一词，在文学领域指的是精神上的无动于衷，思维不敏捷，反应迟钝；在医学领域，则指的是身体上丧失知觉。也可能是一个术语在同一领域却存在多种不同的解释，如"三焦"在中医有 3 种含义，一是指六腑之一，为上、中、下三焦合称；二是指耳穴名；三是推拿部位名。一词多义会使得中医诊疗知识库在术语选择定义时造成困难，容易造成词义上的混乱，理解困难，不利于术语的规范化。因此，对于一词多义的现象，提出中医诊疗知识库规范档建立方法与流程，如图 5 所示：

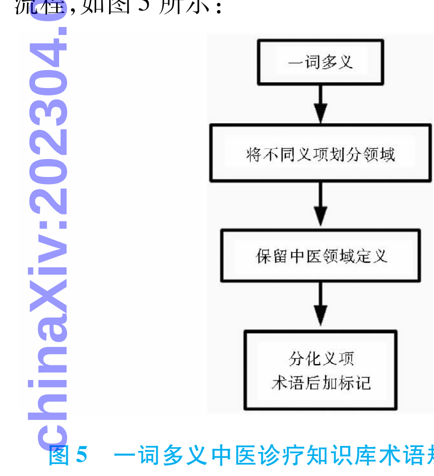


图 5 一词多义中医诊疗知识库术语规范档建立流程

对于一词多义的问题，首先将这一术语的多种定义都列举出，按照领域进行划分。然后判断这些定义是否同属于中医的领域。因为目的在于建立中医诊疗知识库术语规范档，因此可以删除非中医领域的定义项，只保留所需的中医定义。若删除后只剩一个中医定义，可以直接确定该术语的定义。若还存在多个中医定义时，可以将义项分化，通过在术语名词后加上标记的形式，使得一个义项对应一个术语，从而实现一词一义。下面以"三焦"这一中医术语为例。

6.2.2 术语提取及规范档建立——以"三焦"为例

"三焦"有 3 个不同的定义，第一个定义是指人体的六腑之一，为上、中、下三焦的合称；第二个定义是指耳穴名，三焦耳穴，位于外耳门后下，肺和内分泌穴之间；第三个定义是推拿部位的名称，有四个位置，按摩这些部位用于健脾和胃和治疗消食等症状。由于这 3 个定义都属于中医领域，因此需要对不同的义项进行分化。在录入术语名称时，于"三焦"名词后标记上耳穴名，如"三焦(耳穴名)"，就可以对应第二个定义解释。同样，在"三焦"名词后标记上推拿部位名，如"三焦(推拿部位名)"，则对应的是第三个定义。没有加上标注的"三焦"对应的是第一种最为广泛运用的解释。通过这种方式就能够将一词多义转换为一词一义。

6.3 术语概念及关系不明确

6.3.1 建立方法及示意图

术语概念反映的是术语的定义与内涵，术语关系能够体现术语间的层次与逻辑，它们都是构建中医诊疗知识库术语集的重要组成部分。术语概念表达不明确及关系表示混乱，容易造成信息的缺失或冗余，导致用户不能准确获取有效的信息，造成理解上的困难，因此这一问题应予以重视，下面介绍中医诊疗知识库术语概念及关系建立的方法。

(1)概念构成。在设定中医诊疗知识库术语概念的时候，应该遵循中医术语的特点，由该术语的关系特征和属性特征组成。关系特征在此指的是术语的上位词，表示该术语从属于哪一个类、归属于哪一个范围；属性特征描述的是该术语的相关属性及特点。如果说关系特征为纵向概念连接，属性特征则为横向概念连接，共同定义术语的概念。

(2)术语关系建立。中医诊疗知识库中的术语主要划分为五个类别，即：病症名称、症状表现、治疗方式、疗效标准及药品名称，术语参照的标准详见上文 3.4 术语规范档类别与标准的对应关系部分。因此，在构建术语间关系时，采取的方式是以病症名称为原点进行发散，并在以上五个术语类别的基础上，补充病症的上下位词以及它的相对词，相对词指的是同义词、相近词及反义词等。如图 6 所示：

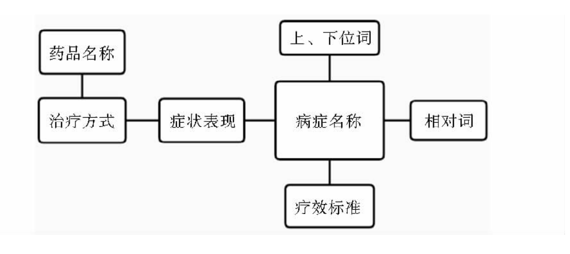


图 6 中医诊疗知识库术语关系建立示意图

6.3.2 术语提取及规范档建立——以"哮病"为例

(1)概念。中医国家标准疾病部分对于"哮病"的定义是：是一种发作性的肺系疾病，通常是由于外邪，

马捷，吴文玲，孙恒宇，等. 中医诊疗知识库术语规范化研究[J]. 图书情报工作,2021,65(2):17－26.

情志损伤,饮食不规律等原因所导致的,临床以突发性呼吸急促困难,喉间哮鸣有声为特征[27]。因此,“哮病”的概念定义可以记为:(上位词)肺系病类;(病因)外邪、情志、饮食;(临床表现)突发呼吸急促困难、喉间哮鸣有声。

(2)术语关系建立。依据上述示意图建立哮病的关系图,如图 7 所示。在此,“哮病”的病症名称、上位词及症状表现参照的是国家标准－中医病症分类与代码及中医临床诊疗术语疾病部分;疗效标准参照的是行业标准－中医内科病证诊断疗效标准;治疗方式参照的是国家标准－中医临床诊疗术语治法部分;药品名称参照的是《中医内科学》第二章肺病证的第二节哮病,这里的药品名称只是“哮病”治疗一般使用的药物,不做药方推荐,具体需要服用的药物需根据自身情况,咨询医师。

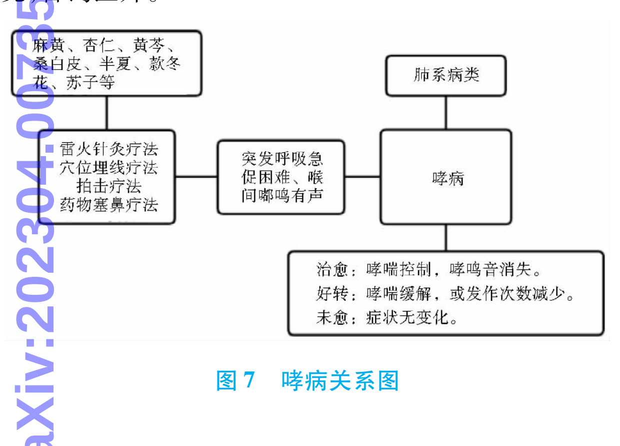


图 7　哮病关系图

7　结语

中医诊疗知识库涉及大量中医术语,术语的规范化工作是构建中医诊疗知识库的基础,也是实现机器理解的前提。笔者基于中医诊疗知识库,确定了中医术语的提取及选词范围,先以已颁布的中医术语国家、行业标准为基本依据,再从权威的中医词典、教科书、核心期刊中选取合适的中医现代用语进行补充。其次,初步制定了中医诊疗知识库的基本结构及术语格式规范,阐明了诊疗知识库的术语规范档类别及其与标准的对应关系。再次,列举了中医诊疗知识库术语规范化存在多词一义、一词多义、用词习惯不统一、病名生僻难懂、术语概念及关系不明确的问题,并对此提出了词义控制策略。最后,以多词一义、一词多义为例,结合“营气”“荣气”“湿热蕴脾”“脾经湿热”术语概念及关系不明确、“三焦”“哮病”的具体实例,论述了中医诊疗知识库术语提取及规范档建立的过程。

中医术语规范化对于弘扬中医优秀文化、推广中医学知识、推动中医现代化与国际化发展具有深远的意义[28]。但是由于中医术语数量多导致术语规范化工作量大,并且需要在实际临床诊疗中反复多加修改。笔者仅对于中医诊疗知识库中的术语规范化问题提出初步的构想,考虑的问题不够全面,在规范化的过程中可能存在疏漏和不合理的地方,需要在实践中加以完善。

参考文献:

[1] 国务院. 国务院印发《中医药发展战略规划纲要(2016－2030年)》[N]. 人民日报,2016－02－27(1).

[2] 邱均平,陈敬全. 知识仓库及其在企业管理中的应用[J]. 情报理论与实践,2003(4):324－326,323.

[3] 朱建平. 中医术语规范化与中医现代化国际化[J]. 中华中医药杂志,2006(1):6－8.

[4] 董燕,李海燕,崔蒙,等. 中医临床术语系统建设概况与改进措施[J]. 医学信息学杂志,2014,35(8):43－48.

[5] 崔蒙,李海燕,杨硕,等. 中医药信息学理论科学领域研究进展[J]. 中国中医药图书情报杂志,2015,39(5):1－6.

[6] 郭玉峰,谢琪,周霞继,等. 构建中医临床术语标准真实世界规范化应用技术体系的思考[J]. 中医杂志,2015,56(7):557－561.

[7] 倪菲,曲金桥,王彩霞,等. 基于文献计量学方法的中医脾胃病症状术语标准化研究[J]. 辽宁中医杂志,2018,45(10):2086－2087.

[8] 李灵巧,包素珍. 论中医术语标准化[J]. 陕西中医学院学报,2015,38(4):14－16.

[9] 朱建平. 中医药名词术语规范化的实践与思考[J]. 中国科技术语,2017,19(6):11－14.

[10] 许吉,施毅,袁敏,等. 中医术语国家标准比较研究[J]. 时珍国医国药,2015,26(9):2294－2295.

[11] 王传池,杨燕,胡镜清,等. 中医药标准化调研分析的现状、问题与相关政策解读[J]. 中华中医药杂志,2018,33(9):3776－3780.

[12] 张悦悦,许鑫. 基于主题词表的亚健康人群食疗知识库设计与构建[J]. 图书情报工作,2015,59(14):6－14.

[13] 马钰,王宏志,高宏. 基于医疗知识库的辅助诊疗系统[J]. 计算机应用,2016,36(S1):217－219,261.

[14] 洪伟. 术语及术语标准的审核与编辑加工要点[J]. 中国质量与标准导报,2017(8):44－46.

[15] 傅柱,王曰芬. 共词分析中术语收集阶段的若干问题研究[J]. 情报学报,2016,35(7):704－713.

[16] 国家技术监督局. GB/T 16751.1－1997 中医临床诊疗术语疾病部分[S]. 北京:中国标准出版社, 1997.

[17] 国家中医药管理局. ZY/T001.5－1994 中医眼科病证诊断疗效标准[S]. 北京:中国标准出版社,1995.

[18] 王刘安,常春. 同义术语归并中缩略语的处理方法研究[J]. 图书情报工作,2014,58(2):121－125.

[19] 常春. 基于词频信息确定叙词表概念属性[J]. 图书情报工作,

2013,57(16):11-14,24.
[20] 苏礼. 一词多义(一)[J]. 陕西中医,1981(2):48.
[21] 许志泉. 中医学术语的多义性及标准化[J]. 山东中医学院学报,1994,18(5):329.
[22] 董兴鲁,曹克刚,马斌,等. 大数据时代下真实世界中医术语研究浅析[J]. 中华中医药杂志,2014,29(9):2724-2726.
[23] 中医药学名词审定委员会. 中医药学名词(2013)[M]. 北京:科学出版社,2014.
[24] 刘旺华,李花,周小青. 关于中医病名规范化的思考——学习《中医临床诊疗术语》有感[J]. 光明中医,2002(1):5-7.
[25] 常春. 从"牛腿"谈概念术语的选择与词间关系建立[J]. 情报杂志,2009,28(10):24-27.
[26] 国家技术监督局. GB/T 16751.2-1997 中医临床诊疗术语证候部分[S]. 北京:中国标准出版社, 1997.
[27] 国家技术监督局. GB/T 16751.1-1997 中医临床诊疗术语疾病部分[S]. 北京:中国标准出版社, 1997.
[28] 张盼,毛树松,邓文萍. 中医药信息标准术语规范化研究[J]. 医学信息学杂志,2018,39(5):12-15.

作者贡献说明:
马捷:提出研究命题、论文整体框架,论文修改及定稿;
吴文玲:负责论文撰写与修改;
孙恒宇:负责论文撰写与修改;
王春秀:负责论文撰写与修改。

Study on the Standardization of Terms in TCM Diagnosis and Treatment Knowledge Base

Ma Jie[1,2] Wu Wenling[1] Sun Hengyu[3] Wang Chunxiu[1]
[1] School of Management, Jilin University, Changchun 130022
[2] Information Resources Research Center, Jilin University, Changchun 130022
[3] The First Clinical Hospital of Jilin Academy of Traditional Chinese Medicine, Changchun 130022

Abstract: [**Purpose/significance**] TCM terminology is the carrier of TCM knowledge and an important part of TCM diagnosis and treatment knowledge base. The research on the standardization of TCM terminology for TCM diagnosis and treatment knowledge base is conducive to improving the quality of medical case data information in TCM diagnosis and treatment knowledge base and promoting TCM culture. [**Method/process**] This paper determined the scope of terms extraction and selection in TCM diagnosis and treatment knowledge base, constructed the structure of terms standard file, and put forward the strategy of word meaning control according to the problems existing in terms standard, and explained the process of terms extraction and standard file establishment with specific example. [**Result/conclusion**] The standardization of terms based on TCM diagnosis and treatment knowledge base can greatly improve the utilization efficiency of TCM diagnosis and treatment knowledge base, and promote the modernization and internationalization of TCM.

Keywords: TCM diagnosis and treatment knowledge base terminology of traditional Chinese Medicine standardization of terms